CATEGORICAL DATA ANALYSIS II: Confounding and Stratified Data Analysis

Categorical Tests of Association

- **Strength of Association**
  - OR – 2x2 Tables
  - Spearman – RxC Tables

- **Statistical Tests**
- **Chi Square – Nominal x Nominal**
  - Uncertainty Coefficient for Strength
- **M-H Chi Square - Ordinal X Ordinal**
  - Spearman for strength
- **Pearson Chi Square with Exact– small numbers**
  - Fisher's – 2x2 Tables

Brief Review of Confounding

- a cause of the disease under study
- associated with exposure
- not a mediator
  - an extraneous variable that wholly or partially accounts for the apparent effect of exposure on disease (Schlesselman)
  - Any cause (risk factor) for a disease is a potential confounder
  - may be a ‘proxy’ for a cause and association independent of exposure (Hennekens)
  - positive or negative confounding

Breast Feeding and Breast Cancer

- more breast feeding = less breast CA
  - *Breast feed to protect against cancer….*
- more breast feeding = more children
  - *Have babies to protect against cancer…*
- more children = young age at first full-term pregnancy
  - *Have babies early and often…*

Comparing the unadjusted to the adjusted measure of effect is the key to assessing for confounding

Myocardial Infarction				
Coffee Drinking		Yes	No	
	Yes	90	60	150
	No	60	90	150
		150	150	300

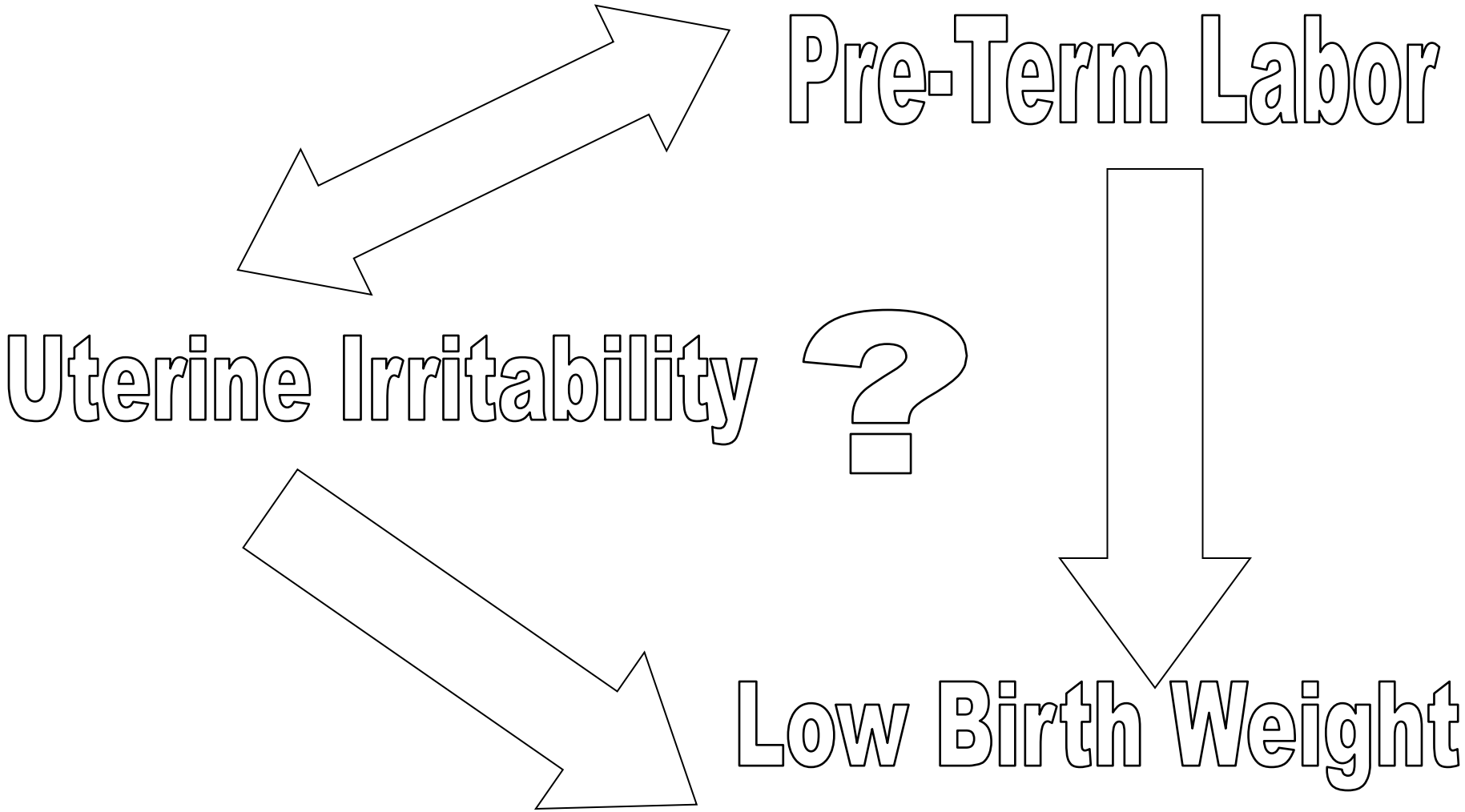
Coffee Drinking and MI:

- OR = ad/bc = (90)(60) / (60)(90) = 2.25
- Smoking?
- Look at data within homogenous groups with and without the potential confounder

	Smokers				Non-Smokers	
	MI	No MI		MI	No MI	
Coffee	80	40	120	10	20	30
No Coffee	20	10	30	40	80	120
Totals	100	50		50	100	
	OR = 1.0			OR = 1.0		

Controlling for Confounding

- **Design**
  - **randomization**
  - **restriction**
  - **matching**
- ***Analysis***
  - ***Stratification***
    - ***compare the crude to the adjusted measure of effect***
  - **multivariate analysis**


Pre-Term Labor
Uterine Irritability
?
Low Birth Weight

Cochran-Mantel-Haenszel statistic (CMH)

- **combines the results of the stratum-specific comparisons**
  - **weighted average of the stratum effects**
  - **weight based on the precision of the effect and the size of the stratum**
- ➔ ***results in an adjusted OR***
- ***large + association at one level, can affect a - association at another***
  - ***Consider interaction first***
- **3 types of CMH:**
  - **Type 1 for linear association of ordinal by ordinal**
  - **Type 2 that raw mean scores of ordinal differ by nominal**
  - **Type 3 for general association of nominal by nominal**

Cochran-Mantel-Haenszel Statistics

Row Variable	Column Variable	CMH Type	Alternate Hypothesis
Ordinal	Ordinal	1	Linear Association
Nominal	Ordinal	2	Row Mean Scores Differ
Nominal	Nominal	3	General Association

Adjusted ORs in PROC FREQ

- the Mantel-Haenszel (MH) estimator
  - weighted average of stratum specific OR's
  - can handle zero frequencies
- logit-based estimator
  - weighted average of log-odds ratios
  - zero frequencies a computational problem
  - adds ½ to zero cells
- Recommendation to use MH estimator with small sample sizes

Adjusted Odds Ratio

Mantel-Haenszel estimator

- weighted average of stratum-specific odds ratios
- handles zero frequencies with no difficulty.

Logit-based estimator

- weighted average of stratum-specific log-odds ratios
- has a problem with zero frequencies.

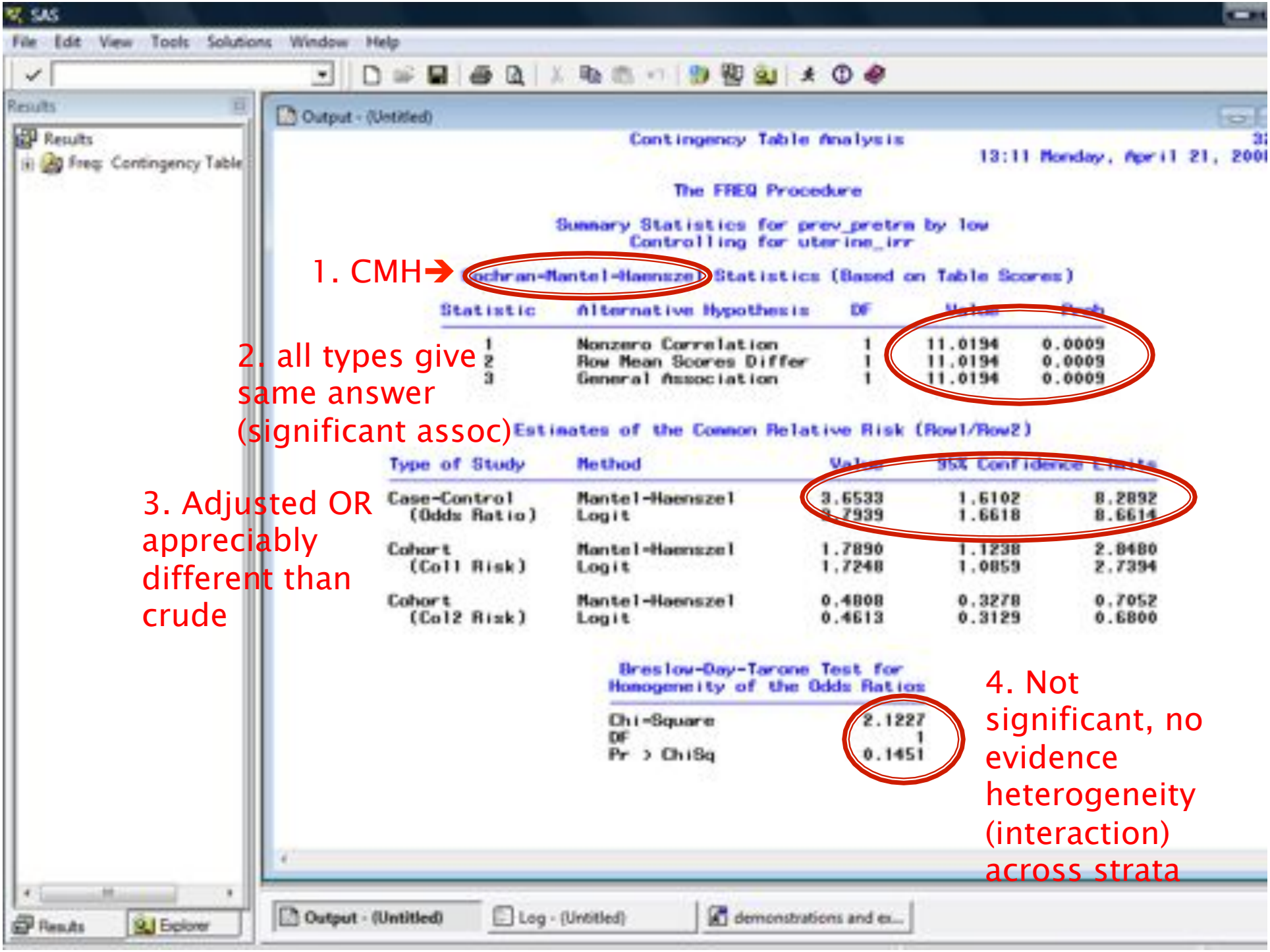
Confounding in the Setting of Interaction

- MH based on assumption OR constant across strata
  - assumption is not met in the setting of interaction
  - *Important to check for interaction first.*
- Breslow-Day statistic
  - check for homogenieity of OR across strata
  - has $\chi 2$ distribution
  - requires appropriately large sample size (recall, 80% cells > 5, no cells < 1)
    - *Tarone's adjustment to correct for a perceived inefficiency of $\chi 2$ when small sample sizes*
    - *SAS also has options to request exact tests for odds ratios when small numbers*
  - can be misleading if OR's vary across 1
- Requesting B-D also provides the CMH

Stratified Analysis Syntax

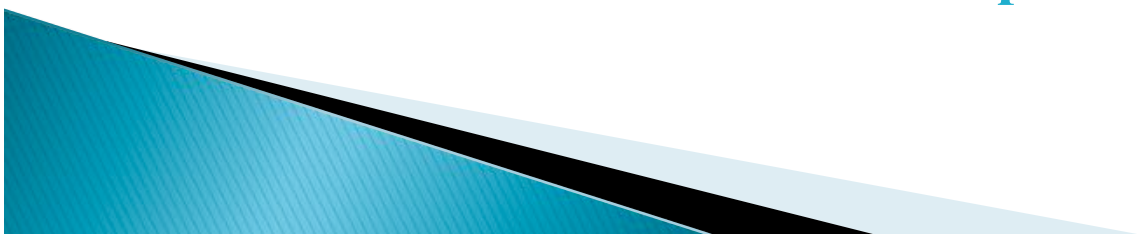
- stratum-specific estimates:
  - proc freq data=your.data;
  - tables confounder*exposure*outcome /
  - measures;
  - run;

- pooled estimate
  - proc freq data=your.data;
  - tables confounder*exposure*outcome /
  - all bdt;
  - run;


SAS
File Edit View Tools Solutions Window Help
Results
Results
Freq: Contingency Table
Output - (Untitled)
Contingency Table Analysis
13:11 Monday, April 21, 200
The FREQ Procedure
Summary Statistics for prev_pretrm by low
Controlling for uterine_irr
1. CMH➔
Cochran-Mantel-Haenszel Statistics (Based on Table Scores)
Statistic Alternative Hypothesis DF Value Prob
1 Nonzero Correlation 1 11.0194 0.0009
2 Row Mean Scores Differ 1 11.0194 0.0009
3 General Association 1 11.0194 0.0009
2. all types give same answer (significant assoc)
Estimates of the Common Relative Risk (Row1/Row2)
Type of Study Method Value 95% Confidence Limits
Case-Control (Odds Ratio) Mantel-Haenszel 3.6533 1.6102 8.2892
Logit 3.7939 1.6618 8.6614
Cohort (Col1 Risk) Mantel-Haenszel 1.7890 1.1238 2.8480
Logit 1.7248 1.0859 2.7394
Cohort (Col2 Risk) Mantel-Haenszel 0.4808 0.3278 0.7052
Logit 0.4613 0.3129 0.6800
3. Adjusted OR appreciably different than crude
Breslow-Day-Tarone Test for Homogeneity of the Odds Ratios
Chi-Square 2.1227
DF 1
Pr > ChiSq 0.1451
4. Not significant, no evidence heterogeneity (interaction) across strata
Output - (Untitled)
Log - (Untitled)
demonstrations and ex...
Results
Explorer

Stratified Analysis in SAS

- PROC FREQ for univariate and stratified analysis
- for 2x2 table OR is the measure of choice, if > 2x2, other options
- Key to looking for confounding and interaction, is to compare the crude to the adjusted rates
- when you find confounding, consider interaction
  - breslow-day to assess homogeneity of OR
- do frequency table analysis for all potential confounders
  - create a table comparing crude effects to adjusted effects
  - summarize main effects and potential confounders



What about 2x2 tables….?

Outcome				
Treatment		Favorable	Un-favorable	
	Placebo	16	48	64
	Test	40	20	60

What about 2x2 tables…?

```
data respire;
   input treat $ outcome $ count ;
   cards;
   placebo f 16
   placebo u 48
   test    f 40
   test    u 20
   ;
proc freq;
   weight count;
   tables treat*outcome;
run;

/* pages 11-12 of Categorical Data Analysis */
```

DEMONSTRATION: STRATIFIED ANALYSIS AND CONFOUNDING

PUTTING IT TOGETHER: CATEGORICAL II

ASSIGNMENT 2: CATEGORICAL TESTS OF SIGNFICANCE, MEASURES OF ASSOCIATION, CONTINGENCY TABLE ANALYSIS